

【目的】 タンパク質とリガンド（低分子化合物）の結合親和性（affinity）の解析は、薬剤および有用酵素の設計において、基盤となる重要な技術である。結合親和性をもとに、特定のタンパク質に結合するリガンドの探索、あるいは、特定のリガンドに結合するタンパク質の探索が行われるが、こうしたスクリーニングを広範に行うには、生化学実験や構造解析は多くの時間と労力を必要とするため、生命情報工学による予測が期待されている。本研究では、タンパク質とリガンドが結合する可能性のある部位の 3 次元（3D）構造から、深層学習によって、それらの結合親和性を高精度で予測（回帰）するシステムを開発する。実験によって得られた結合構造を学習させ、ドッキング予測などで得られた候補構造に対して予測を行う。予測には、結合構造あるいは候補構造の周辺の 3D 空間をグリッドに区切り、グリッド内の原子の属性（原子種など）を入力とする。このような予測には高度な深層学習が不可欠であり、本研究では、画像認識の分野で実績のある convolutional neural network（CNN）を 3D 空間に適用できるよう拡張した手法 3D-CNN を開発する。

【方法】 1. 学習データセットの取得：学習に用いるデータは、タンパク質-リガンドの結合構造と実験によって計測された親和性の値を収集した PDBbind データベース (<http://www.pdbbind.org.cn/>) から取得した。性能評価は、PDBbind のデータのうち、信頼度の高いものを選択して行った。2. 入力データを処理するプログラムの開発：タンパク質-リガンドの結合構造を深層学習に輸入するため、タンパク質-リガンド結合部位の空間を 3 次元ボクセルに分割し、各ボクセルに位置するタンパク質、リガンドそれぞれの構成原子に対し、原子種とその構造的、化学的環境に応じて原子のタイプを定義し、チャンネルに割り当てた。3. 深層学習による予測プログラムの開発：タンパク質-リガンド結合部位の周辺を直接学習させるため、3D-CNN を開発し、タンパク質の特性と合わせて結合親和性を学習するシステムを構築した。また、構造と合わせて、PSI-BLAST を用いて類縁タンパク質のマルチプルアラインメントを構築し、プロファイルを作成した。プロファイルはネットワークの最後のステージに輸入することで、進化情報を加味した予測を行った。4. テスト予測とパラメータの調整：PDBbind から選択した学習データセットを対象にテスト予測を行い、3D-CNN の畳み込み層およびプーリング層の構成、フィルターの設計、その他各種パラメータの最適化を行った。

【結果】 PDBbind のデータに対して予測した親和性と実験によって得られた親和性（どちらも pK 値）の関係を示したものである。相関係数は 0.782 であり、3D 構造と従来の配列ファイルのみを用いた比較的簡易な方法で高い予測性能が得られていることがわかる。

結合親和性予測の結果と実験値との比較

